New Phytologist

Optimizing the statistical estimation of the parameters of the Farquhar–von Caemmerer–Berry model of photosynthesis

Jean-Jacques B. Dubois[1,2], Edwin L. Fiscus[1,2], Fitzgerald L. Booker[1,2], Michael D. Flowers[1,3] and Chantal D. Reid[1,4]

[1]United States Department of Agriculture, Agricultural Research Service, Plant Science Research Unit, 3127 Ligon Street, Raleigh, NC 27607, USA; [2]Crop Science Department, North Carolina State University, 3127 Ligon Street, Raleigh, NC 27607, USA, [3]Department of Crop and Soil Science, Oregon State University, 109 Crop Science Building, Corvallis, OR 97331-3002, USA; [4]Department of Biology, Box 90338, Duke University, Durham, NC 27708, USA

Author for correspondence:
Jean-Jacques B. Dubois
Tel: +1 (919) 515 9496
Fax: +1 (919) 515 3593
Email: Jean-Jacques.Dubois@ars.usda.gov

Received: *11 April 2007*
Accepted: *6 June 2007*

Summary

• The model of Farquhar, von Caemmerer and Berry is the standard in relating photosynthetic carbon assimilation and concentration of intercellular CO_2. The techniques used in collecting the data from which its parameters are estimated have been the object of extensive optimization, but the statistical aspects of estimation have not received the same attention.

• The model segments assimilation into three regions, each modeled by a distinct function. Three parameters of the model, namely the maximum rate of Rubisco carboxylation ($V_{c\,max}$), the rate of electron transport (J), and nonphotorespiratory CO_2 evolution (R_d), are customarily estimated from gas exchange data through separate fitting of the component functions corresponding to the first two segments. This disjunct approach is problematic in requiring preliminary arbitrary subsetting of data into sets believed to correspond to each region.

• It is shown how multiple segments can be estimated simultaneously, using the entire data set, without predetermination of transitions by the investigator.

• Investigation of the number of parameters that can be estimated in the two-segment model suggests that, under some conditions, it is possible to estimate four or even five parameters, but that only $V_{c\,max}$, J, and R_d, have good statistical properties. Practical difficulties and their solutions are reviewed, and software programs are provided.

Key words: carbon assimilation, maximum rate of Rubisco carboxylation ($V_{c\,max}$), nonlinear models, nonphotorespiratory CO_2 evolution (R_d), photosynthesis model, rate of electron transport (J), segmented regression.

New Phytologist (2007) **176**: 402–414

doi: 10.1111/j.1469-8137.2007.02182.x

Introduction

Since its publication in 1980, the model of Farquhar, von Caemmerer and Berry (FvCB; see Table 1 for list of abbreviations) has become the standard in understanding and quantifying the kinetics of carbon fixation by photosynthesis in terrestrial plants (Farquhar *et al.*, 1980; von Caemmerer, 2000). For C_3 plants, the model, in its simplest form, summarizes the dependence of carbon assimilation rate (A) on intercellular CO_2 partial pressure (C_i) as determined by saturation of Rubisco with respect to carboxylation, electron transport as limited by ribulose bisphosphate (RuBP) regeneration, or triose phosphate export. At any given partial pressure of CO_2, A is modeled as the smallest value of the corresponding three functions, A_c, A_j, and A_p, adjusted for nonphotorespiratory CO_2 release, R_d (Fig. 2). In the currently prevailing notation:

Table 1 Definition of abbreviations

Abbreviation	Definition	Units
A	Assimilation rate	μmol m^{-2} s^{-1}
A_c	Rubisco carboxylation-limited (i.e. RuBP-unsaturated) assimilation rate	μmol m^{-2} s^{-1}
A_j	RuBP regeneration-limited assimilation rate	μmol m^{-2} s^{-1}
A_p	TPU-limited assimilation rate	μmol m^{-2} s^{-1}
$V_{c\,max}$	Maximum rate of Rubisco carboxylation	μmol m^{-2} s^{-1}
J	Rate of electron transport	μmol e^- m^{-2} s^{-1}
C_c	CO_2 partial pressure at the site of carboxylation	μmol mol^{-1} or μbar
C_i	Intercellular CO_2 partial pressure	μmol mol^{-1} or μbar
$C_{i\,tr}$	Value of C_i at the transition point between A_c and A_j	μmol mol^{-1} or μbar
Γ_*	Photosynthetic compensation point	μmol mol^{-1} or μbar
R_d	Nonphotorespiratory CO_2 evolution	μmol m^{-2} s^{-1}
K_c	Michaelis–Menten constant of Rubisco for CO_2	μmol mol^{-1} or μbar
K_o	Michaelis–Menten constant of Rubisco for O_2	mmol mol^{-1} or mbar
O	Partial pressure of O_2	mmol mol^{-1} or mbar
T_p	Rate of TPU	μmol m^{-2} s^{-1}
α	Nonreturned fraction of glycolate	μmol m^{-2} s^{-1}
FvCB model	The Farquhar–von Caemmerer–Berry model of the response of carbon assimilation to CO_2 concentration	

RuBP, ribulose bisphosphate; TPU, triose phosphate utilization.

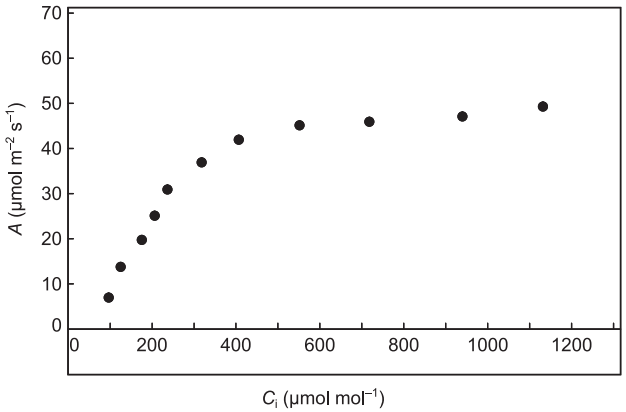

Fig. 1 Typical set of A/C_i data (response of photosynthetic assimilation (A) to varying intercellular partial pressure of CO_2 (C_i)) for soybean 'Essex' (*Glycine max* 'Essex').

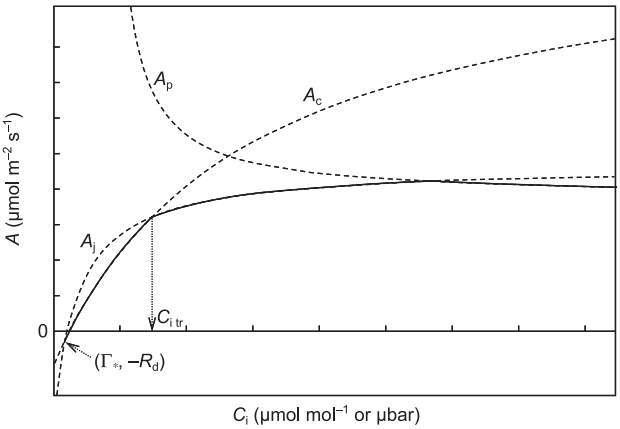

Fig. 2 Photosynthetic assimilation (A) as a function of intercellular partial pressure of CO_2 (C_i). A_c, A_j, and A_p are three assimilation functions (as described in the text and Table 1). The solid line is A, as modeled by the Farquhar, von Caemmerer and Berry (FvCB) model.

$$A = \min\{A_c, A_j, A_p\} \quad \text{Eqn 1}$$

with

$$A_c = \frac{V_{c\,max}(C_i - \Gamma_*)}{C_i + K_c\left(1 + \frac{O}{K_o}\right)} - R_d \quad \text{Eqn 2}$$

$$A_j = \frac{J(C_i - \Gamma_*)}{4C_i + 8\Gamma_*} - R_d \quad \text{Eqn 3}$$

$$A_p = \frac{3T_p(C_i - \Gamma_*)}{C_i - (1 + 1.5\alpha)\Gamma_*} - R_d \quad \text{Eqn 4}$$

Sets of measurements can be obtained with relative ease, in which A is recorded at varying values of C_i. In a majority of reports, eight to 12 measurements are taken on one leaf (or one branch), with eight to 12 values of C_i from 50 to 1500 μmol mol^{-1}, or some narrower range. An example is presented in Fig. 1. We will refer to such a set, consisting of a series of (C_i, A) points, as an 'A/C_i set'. In consideration of the great rarity of studies where triose phosphate-limited data have been presented, the discussion will focus on the two-segment model (A_c and A_j), with a few applicable notes on the three-segment model.

Segments defined by the functions A_c, A_j, and A_p combine to form the continuous assimilation function A if, and only if, they intersect with one another. A_c and A_j intersect twice (Fig. 2), and the ordinate or C_i values of the two intersections

are given by solving $[A_c = A_j]$ for C_i. The first intersection is thus at $C_i = \Gamma*$ (and $A = -R_d$), and we define $C_{i\,tr}$ as the second intersection, given by the other solution:

$$C_{i\,tr} = \frac{K_c J(K_o + O) - 8K_o\Gamma* V_{c\,max}}{K_o(4V_{c\,max} - J)} \qquad \text{Eqn 5}$$

$C_{i\,tr}$, or C_i at the transition between A_c and A_j, is fully defined by the model, and does not introduce any new information or parameter. By definition of the model, this transition between A_c and A_j varies with every A/C_i set.

A_c and A_j are both asymptotic functions of C_i, with $V_{c\,max}$ and J each a direct proportion of the respective horizontal asymptote. The maximum rate of electron transport, J_{max}, in turn, is a direct proportion of the asymptote of J under increasing light. As long as gas exchange is measured under saturating light, J can be assumed equal to J_{max}. If light was not saturating at the time of measurement, J_{max} must be calculated from J, either using the results from fitting a light–response function to data from the same plants, or borrowing a previous parametrization of that function. The coefficients of C_i and $\Gamma*$ in the denominator of Eqn 3 are modified according to whether ATP or NADPH shortage is believed to underlie electron transport limitation. The latter assumption is used most often, corresponding to $4C_i$, and $8\Gamma*$.

Finally, the model was in fact reasoned for the response of A to C_c, or varying CO_2 partial pressure at the chloroplast. C_c is related to C_i, A, and the conductance of the path from the substomatal cavity to chloroplasts (g_i) by the simple equation:

$$C_c = C_i - (A/g_i) \qquad \text{Eqn 6}$$

Because C_c cannot be measured, usage has been to employ C_i, or intercellular CO_2, and in practice the two have thus been assumed to be equal. Problems with this assumption have often been raised, as reviewed in Ethier & Livingston (2004).

The two-segment model includes six parameters: $V_{c\,max}$, J, R_d, $\Gamma*$, K_c, and K_o. Only three are usually computed from A/C_i data: $V_{c\,max}$, J, and R_d. Values for the others are taken instead from some other source, and 'plugged in'. It is important to note that values for $V_{c\,max}$, J, and R_d are not measured, nor are they calculated algebraically from measurements. Instead, they are estimated through fitting the model to sets of measurements. They are not observations, but products of statistical estimation. For clarity, we will adhere to the following terminology: $V_{c\,max}$, J, R_d, $\Gamma*$, K_c, and K_o, are all 'parameters'. When some parameters are not included in the estimation process, and their value is 'plugged in', they will be referred to as 'set' or 'fixed'. This is in contrast with the convention of reserving the term 'parameter' for coefficients that are the object of estimation, and 'constant' for all others. 'Estimation' and 'estimate' are used in their statistical sense, never as synonyms for 'approximation'; for every plant part from which an A/C_i set is collected, there is a true, but unobserved value for the parameters of the model, and the process of fitting the model to those data should yield the best possible estimate of that true value, which remains unobserved.

The validity of the FvCB model has been abundantly corroborated, and it has been incorporated into most crop and ecosystem simulation models. Furthermore, its suitability for describing fundamental responses of photosynthesis is such that $V_{c\,max}$ and J_{max} have become valuable metrics of photosynthetic performance in themselves. The task of computing them from gas exchange data is thus carried out primarily to either parametrize predictive models, or for its own sake, in order for example to characterize the effect of some factor of interest on photosynthesis.

Within the chain of procedures through which values for the parameters of the FvCB model are derived, the instruments and techniques used in measuring A and C_i have been the object of continuous optimization over the last 25 yr. Similarly, the methods used to obtain values for fixed parameters have seen important strides. Both aspects are summarized in Long & Bernacchi (2003), and efforts to optimize them continue to this day (see for example Flexas *et al.*, 2007). In contrast, the methods used in fitting the model, that is, estimating the parameters of interest, have not received as much attention. Treatment of the statistical methods used in estimating parameters from A/C_i data has been both exceedingly rare and brief. The aim of this article is to suggest improvements to that link of the chain.

Multiple segments of models such as the FvCB model can be estimated simultaneously, with adjustments to all estimated parameters performed at once, based on the entire data set. All but a very few of the studies that have estimated the model have not been aware of this possibility, and have instead relied on estimating each segment separately. We review established estimation methods for the FvCB model, describe simultaneous estimation of multiple segments, and briefly provide some details of implementation. We also consider some of the properties of the FvCB model that can only be investigated once a method of estimating more than one segment at a time is available. Implementation requires care in avoiding some potential pitfalls, which are reviewed, along with suggested solutions. Programs, written for the SAS System (SAS Institute, Cary, NC, USA), are presented as Supplementary Material Appendix S1, and can be adapted to other software such as R.

Disjunct segments estimation methods

Fitting the FvCB model to gas exchange data is customarily accomplished by fitting each component function separately (disjunct segments estimation). This requires that each A/C_i set be first divided into two subsets, each believed to comprise samples of one segment or the other. Preliminary subsetting has been implemented in several ways. It is necessarily arbitrary, in the sense of being at the discretion of the investigator.

Subsetting is usually subjective: each A/C_i set is examined individually, and a different subsetting point (cut-off) is chosen for each. The subset of observations below the cut-off is fitted using Eqn 2, with $\Gamma*$, K_c, and K_o fixed, yielding estimates of $V_{c\,max}$ and R_d. The estimated value of R_d is then used as a fixed value in estimating J by fitting Eqn 3 to the subset defined by C_i being greater than the cut-off. It should be noted that the value of $C_{i\,tr}$ that can be calculated from the resulting parameter estimates is not necessarily the same as the cut-off value. The A/C_i set in Fig. 1 suggests a discontinuity in overall curvature (change point): the observer must choose, without knowledge of the curve, the two observations between which the set is to be divided. This type of subsetting may be accomplished by visual inspection of the plotted observations or, alternatively, by first fitting only Eqn 2, often linearized for convenience (Long & Bernacchi, 2003), to the entire A/C_i set. Beginning with the highest values, observations are gradually removed from the set, simultaneously monitoring changes in the coefficient of determination r^2 with each new fitting of A_c. Upon reaching some chosen value of r^2, the observations still remaining are regarded as belonging in the Rubisco-limited range. The estimates of $V_{c\,max}$ and R_d obtained from fitting A_c to the subset thus defined are accepted, and A_j is fitted to the other subset, with R_d fixed. This method is not valid and, unfortunately, does not provide the objectivity that is hoped for. Maximization of r^2 may be supported in considering parameters for inclusion in a model, but it cannot be used to choose data to be included or excluded from estimation. There is also no objective way to parse the proportion of the observed change in r^2 attributable to approach of the true change point. The upward progression of the r^2 statistic toward unity with the withdrawal of observations is inherent to its definition, and because the progression is nonmonotonic, setting a critical value for r^2 cannot provide an objective criterion.

Arbitrary subsetting can be also be accomplished objectively, as long as every set is subsetted blindly. This is normally accomplished using a single cut-off value of C_i, chosen *a priori*. Separate fitting of A_c and A_j to the two subsets then proceeds as in the first method. Examples include, but are not limited to, Harley & Tenhunen (1991), Wullschleger (1993), Cai & Dang (2002), Schultz (2003), and Ethier & Livingston (2004). von Caemmerer & Farquhar (1981) have been cited in recommending a single cut-off at approx. 200–250 μbar partial pressure of CO_2, but in the original paper it is made clear that this value is applicable only to *Phaesolus vulgaris*, and is only tentative. Similarly, Wullschleger (1993) only called on his personal subjective experience. Others then referenced these authors' values.

Whether objective or subjective, arbitrary subsetting of the data is less than optimal for two reasons. First, it does not proceed from the data themselves, and from the fit of the model to them. Being arbitrary, it creates an entry for systematic deviation from the true parameter values. This is far from a trivial consideration, because, as illustrated in Fig. 3, the choice of where to perform the subsetting has a substantial influence on parameter estimates. When performed subjectively, subsetting undermines reproducibility, and provides ready leverage for influencing results. Irrespective of whether this ever actually occurs, and of whether or not it is inadvertent if it does, this is an exceedingly undesirable situation.

Secondly, another consequence of arbitrary subsetting of the data is that each segment is estimated using only partial data. Instead of all parameters being adjusted simultaneously based on all data, and each other, $V_{c\,max}$ and R_d are estimated from one subset, and J from the other. $V_{c\,max}$ and J are proportional to the asymptote of the two hyperbolas, A_c and A_j, and estimation of a parameter of an hyperbola is especially problematic when few or no observations are available in the region of the curve over which it has control (Seber & Wild, 1989; Ratkowsky, 1990). As long as data are available in the flatter region of A, the estimate of J is thus inherently less sensitive to misspecification of the cut-off. In contrast, by definition of the model, data are never available in the asymptotic region of A_c. Therefore, the estimate of $V_{c\,max}$ from disjunct estimation is necessarily more erratic than that of J, as illustrated by Manter & Kerrigan (2004). Where estimation of R_d is concerned, small differences in the other parameters, and in data, can have large proportional effects, as a result of the high slope of both A_c and A_j at the $(\Gamma*, -R_d)$ point. By using information from the Rubisco-limited region only to estimate R_d, less accuracy is achieved than is possible. As this estimate of R_d, based on fitting A_c only, is used as a fixed parameter in estimation of J from the RuBP regeneration-limited region, some avoidable bias (in the statistical sense) may be introduced in the estimate of J.

An additional problem is the potential for artifactual inflation of the relatedness of $V_{c\,max}$ and J_{max}, as a result of the narrowing of the constraints placed on these parameters. The potential for artifactual inflation of the relatedness of $V_{c\,max}$ and J_{max} stems from the structure of the model itself, and is a particular concern when using the universal cut-off method. The proportionality of $V_{c\,max}$ and J_{max} has been observed in pools of estimates (Wullschleger, 1993; Leuning, 1997; Wohlfahrt *et al.*, 1999; Medlyn *et al.*, 2002; Onoda *et al.*, 2005). Taking into account the fact that, as long as light is saturating, $J = J_{max}$, the proportionality can be seen as a property of the model itself, by rearranging Eqn 5:

$$J = V_{c\,max}\frac{4K_o(2\Gamma* + C_{i\,tr})}{K_c(K_o + O) + K_oC_{i\,tr}} \qquad \text{Eqn 7}$$

As defined in Eqn 7, and illustrated for four values of $C_{i\,tr}$ in Fig. 4, $V_{c\,max}$ and J (J_{max} under saturating light) are a function of each other and of $C_{i\,tr}$, for any given values of $\Gamma*$, O, K_c, and K_o. At any single value of $C_{i\,tr}$, the relationship of $V_{c\,max}$ and J is a line. The space in which $(V_{c\,max}, J)$ points are possible (Fig. 4, shaded areas) is bounded by the range of $C_{i\,tr}$. Any restriction on that range, brought about for instance by

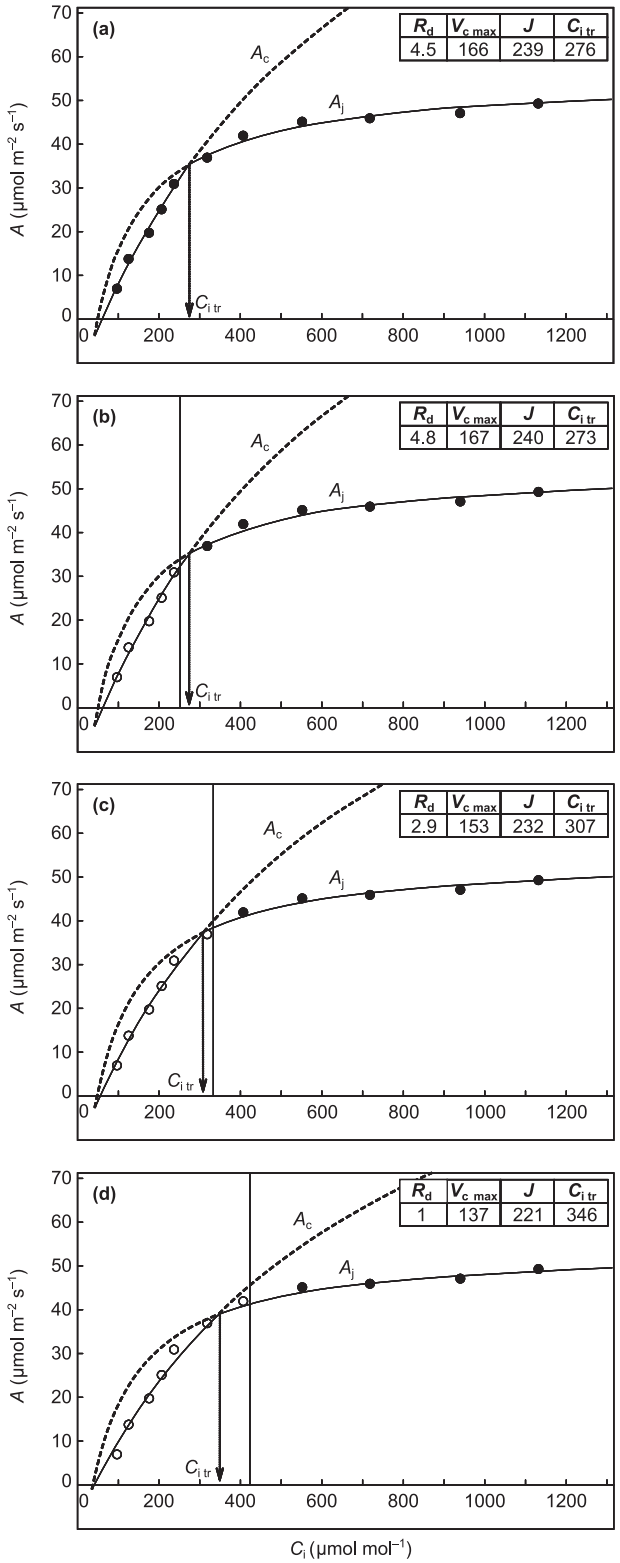

Fig. 3 Effect of alternate choices in preliminary subsetting of A/C_i data (see Table 1 for definition of abbreviations). Data for soybean 'Essex' (*Glycine max* 'Essex') are the same as in Fig. 1. (a) Results from simultaneous estimation. (b–d) Result of changing the cut-off point and performing estimation separately on each subset. A_c was fitted to the lower subset of points, and A_j to the higher. In (b–d), open symbols are used to represent data used to fit A_c, i.e. estimate $V_{c\,max}$ and R_d, and closed symbols for data used to fit A_j, i.e. estimate J with R_d taken from the fitting of A_c. $C_{i\,tr}$ is computed from $V_{c\,max}$, J, and R_d, after they have been estimated. Resultant estimates and fits are shown. Changing subsets affects all parameter estimates, and $C_{i\,tr}$. Subsetting data between the two observed points that bracket the value of $C_{i\,tr}$ obtained from simultaneous estimation (a) leads to the closest results for both methods. Note that $C_{i\,tr}$, the transition point, is different from the cut-off point at which data were subsetted; in (c), $C_{i\,tr}$ is between the fifth and sixth data, as in (a) and (b), even though subsetting was performed between the sixth and seventh.

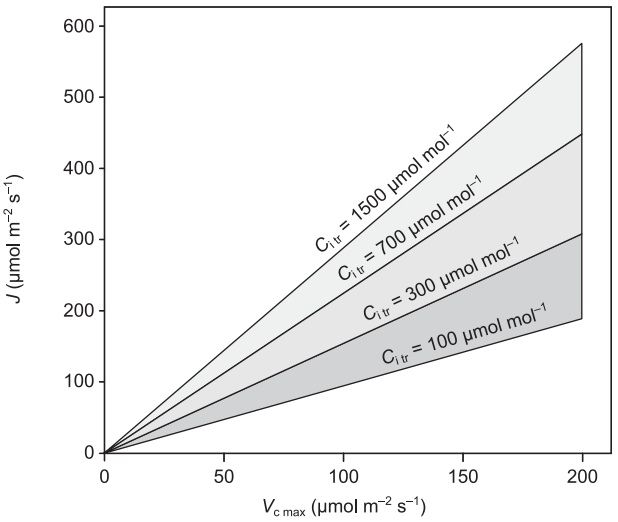

Fig. 4 The role of $C_{i\,tr}$ in defining the relationship of $V_{c\,max}$ and J (see Table 1 for definition of abbreviations). $0 < V_{c\,max} < 200$ μmol m^{-2} s^{-1}; $0 < J < 575$ μmol m^{-2} s^{-1}; $100 < C_{i\,tr} < 1500$ μmol mol^{-1}; $K_c = 405$ μmol mol^{-1}; $K_o = 278$ μmol mol^{-1}; $\Gamma_* = 45$ μmol mol^{-1}; $O = 210$ mmol mol^{-1}. At a single value of $C_{i\,tr}$, the relationship of $V_{c\,max}$ and J is a line. For any range of $C_{i\,tr}$, all possible ($V_{c\,max}$, J) points fall within the shaded area bounded by the corresponding lines.

estimation methodology, restricts the domain of the relationship between $V_{c\,max}$ and J (J_{max} under saturating light). The larger the number of A/C_i sets under consideration, and the more diverse their source, the more likely variability in the ratio of $V_{c\,max}$ and J_{max} is to be underestimated; the less felicitous the cut-off value, the more biased the estimate of that ratio. Underestimation of the variability in the ratio of $V_{c\,max}$ and J_{max} would be reflected in inflated r^2 for the regression of $V_{c\,max}$ on J, while a poor choice of cut-off would affect the slope of that regression. The results of Medlyn *et al.* (2002), having been obtained through simultaneous estimation, are therefore preferable to those of Wullschleger (1993). Interestingly, however, those results suggest that, in practice, this concern may not be warranted. It could be that the biological constraints on $V_{c\,max}$ and J_{max} are already such that constraining $C_{i\,tr}$ only hems in the ($V_{c\,max}$, J) point in a very small proportion of A/C_i sets.

Finally, when using disjunct segment estimation, we can ensure that the fit of A_c and A_j to the respective subsets chosen to estimate them is optimal. The overall fit of the entire model may also be quantified, but we cannot be sure that it is optimal.

More generally, it has not been possible to characterize the estimation behavior of the model, and, in particular, to study how many parameters are in fact estimable.

Simultaneous estimation method

The conditional submodels of the FvCB model (A_c, A_j, and A_p) can all be fitted simultaneously, generating parameter estimates from the entire data set at once. While disjunct approaches never fit more than one segment at a time, simultaneous estimation, sometimes referred to as 'segmented regression', provides the best objective joint fit of multiple segments to data (Seber & Wild, 1989; Rawlings *et al.*, 1998). Simultaneous estimation of the FvCB model can be accomplished readily by any statistical software with the capability to perform nonlinear regression, and to process conditional syntax. At least three studies have made use of simultaneous estimation for the FvCB model, but, to our knowledge, no discussion of this method has been published (Dreyer *et al.*, 2001; Medlyn *et al.*, 2002; Nippert *et al.*, 2007; P. Montpied & B. Medlyn, pers. comm.).

Estimation of nonlinear models is accomplished by minimizing an objective function, using a minimization algorithm. Both elements are chosen independently of one another. Regression methods may differ in the objective function they seek to minimize: ordinary least squares and nonlinear ordinary least squares minimize the error sum of squares (SSE), while maximum likelihood, seemingly unrelated regression, and other methods each minimize a distinct corresponding objective function. In nonlinear models, as opposed to linear ones, minima cannot be found analytically. Nonlinear regression therefore uses iterative algorithms such as Gauss–Newton, steepest descent, or Levenberg–Marquardt algorithms. The value of the objective function is returned for successive combinations of values of the parameters being estimated, until it converges to a minimum, where parameter values are accepted (Seber & Wild, 1989; SAS Institute Inc., 2005). As long as the model is continuous, segmented nonlinear models can be estimated through this process in the same way as other nonlinear models. The only additional requirement is that the software have the capability to process the conditional statements that are required, into a form that is tractable to the minimization algorithm. Most current statistical software has this capability, although not all provide diagnostics and other important options. We limit our discussion to the two-segment model (A_c and A_j), but, should data be encountered that are thought to include data from the triose phosphate utilization (TPU)-limited segment (A_p), this approach can be extended to include it, and thus test for its presence.

Programming

Segmentation can be achieved through a variety of programming syntaxes. Segmented models are characterized by the presence of conditional submodels: if the independent variable is smaller than the join point, the response takes one functional form, and another if it is larger. Alternatively, the overall segmented function takes on the value of one function if its value is smaller than the other at a given value of the independent variable, and of the other function otherwise.

Programming examples in the text are generic, and full programs, written for the SAS System, are presented as Appendix S1. Programming statements are denoted by the font `Courier New`. In the following, '`rubisco`' stands for Eqn 2, and '`RuBP`' for Eqn 3. It is assumed that the values of $V_{c\,max}$, J, and R_d are being estimated, and that K_c, K_o, and $\Gamma*$ are set. Depending on the software used, an explicit expression may or may not be required for $C_{i\,tr}$, the upper join point. The practical consequence of avoiding an explicit expression for $C_{i\,tr}$ is that this allows the procedure to modify the set values of K_c, K_o, and $\Gamma*$ associated with each data point. When fluctuations in temperature were recorded within the course of the A/C_i set, this allows temperature adjustments to the fixed parameters within the procedure. Two forms are possible:

```
...
if rubisco < RuBP then A = rubisco;
else A = RuBP;
...
```

or the equivalent:

```
...
A = min(rubisco, RuBP);
...
```

If software does not permit the adjustment of set parameters within the procedure, $C_{i\,tr}$ must first be defined explicitly (using Eqn 5), and the following syntax may then be used instead:

```
...
if Ci < Citr then A = rubisco;
else A = RuBP;
...
```

This requires that the values of fixed parameters be kept constant for all points within a single A/C_i set, for example by averaging temperature over the course of measurements, but all three options produce identical parameter estimates and statistics.

Differences between disjunct and simultaneous estimates

Despite their evident relatedness, the models estimated by disjunct and simultaneous estimation are different, in that the

former fits two separate models to two separate data sets, while the latter fits a single model to one data set. Parameter estimates will therefore differ, but the magnitude of the difference is expected to vary widely. Insofar as simultaneous estimation fits the more complete model, and uses more information and statistically sound methods, its estimates are expected to be more reliable. They are also entirely reproducible. However, that is not to say that estimates obtained by the two methods cannot be very close to one another, on a biologically meaningful scale.

The magnitude of the difference in parameter estimates depends on both data and observer, and is thus unpredictable. The following reasoning may help frame the question of comparing methods. The value of $C_{i\,tr}$ computed from parameters estimated through simultaneous estimation is the best statistical estimate of the true, but unknown change point between A_c and A_j, or the point of discontinuity in curvature (Seber & Wild, 1989; Rawlings *et al.*, 1998). If an A/C_i set is subsetted between the two observations that bracket the value of $C_{i\,tr}$ estimated by simultaneous fitting, and disjunct estimation is used on those two subsets, estimates of $V_{c\,max}$ and J obtained by the two methods are expected to produce the closest possible agreement between the methods. However, accurate subjective identification of the change point is often precarious. The presence of distinct segments is often far from evident (Harley & Tenhunen, 1991), and even when it is discernible from the data, the location of the change point generally is not. The magnitude of the difference between methods depends both on data and on how far from $C_{i\,tr}$ the observer places the cut-off. Different data and observers will produce varying degrees of agreement between methods, and the more difficult the change point is to identify, the more likely results are to diverge. The data in Figs 1 and 3 are a clear instance of how close the estimates obtained by the two methods can be, if the cut-off for disjunct estimation is chosen close to the estimate of $C_{i\,tr}$ obtained by simultaneous estimation. They also illustrate the impact the choice of subsetting point has on estimates derived by disjunct estimation. Our informal experience with this A/C_i set is that, *without knowledge of the various fits*, some observers place the cut-off point after the fifth observation, some after the sixth, and very few after the seventh. Other sets are more challenging, some less.

It is likely that, for many of the published studies that include estimates of $V_{c\,max}$ and J_{max}, the change that would result from using simultaneous estimation would be minor, but this can only be assessed by processing each data set concerned. It is also likely that, for at least some published studies, the difference between previous estimates and simultaneous ones would not be inconsequential. However, a survey is outside the scope of this article. More generally, an adequate test of either method would entail applying it to a large number of simulated A/C_i sets. In simulated data, the true value of the parameters is known, as it is controlled when generating the data, as is the nature of the error, or noise, that is added to them. The performance of each estimation procedure is measured by its success in retrieving those true values, under various configurations of both parameters and noise, with more than a thousand data sets for each configuration (D. Dickey, pers. comm.). Applying this approach to disjunct estimation of the FvCB model is probably unfeasible, considering that subsetting would have to be performed on every A/C_i set by several independent observers. In addition, it is uncertain whether enough is known of the correlations among parameters in real gas exchange data to generate an accurate representation of real-world data.

Some statistical properties of the FvCB model

With the ability to estimate the whole model comes the possibility of characterizing some of its estimation properties. We first address limits to the number of estimable parameters, in an attempt to determine how many of the six parameters of the model can be estimated reliably from A/C_i data. We then look for the presence of local minima in the objective function, to establish whether the FvCB model is subject to that most common difficulty in fitting nonlinear models.

Estimable parameters

For linear models with k levels of an independent variable, up to $k-1$ parameters can be estimated. This property does not extend to nonlinear models. The number of parameters that can be estimated in nonlinear models, given some number of discrete levels of the independent variable(s), is difficult or impossible to predict analytically. Should an A/C_i set include enough levels of C_i, it is not impossible that all six parameters might become estimable, but gauging how many are, in reality, is difficult. Joint estimation of $\Gamma*$, K_c, and K_o along with $V_{c\,max}$, J, and R_d warrants investigation because, as Ethier & Livingston (2004) have pointed out, the suitability of using generic values for those parameters is not settled, and the value of the fixed parameters has a strong impact on estimates of the parameters of interest. We approached the question both empirically and using Hougaard's measure of skewness. Hougaard's measure of skewness, or g_{1i}, is a statistic that may be computed for each parameter, as it is estimated from a specific data set. It changes with every data set, and with the addition of other parameters to the model. It is used to assess the closeness of individual parameters to linearity (Ratkowsky, 1990; Haines *et al.*, 2004): a close-to-linear parameter is one whose statistical reliability is close to that of parameters of linear models. An estimator is very close to linear if $|g_{1i}| < 0.1$, and reasonably so if $0.1 < |g_{1i}| < 0.15$, but strongly skewed if $|g_{1i}| > 0.15$ (Haines *et al.*, 2004), making inference all but invalid. Its utility is twofold. First, when estimation fails in some proportion of data sets, Hougaard's measure for the successful sets helps to identify the parameter or parameters whose removal is most likely to make estimation possible.

Secondly, it provides an indication of the chance that the true value of a parameter may be captured at all.

Two collections of A/C_i sets were processed using simultaneous estimation, with varying numbers and combinations of parameters to be estimated, among $V_{c\,max}$, J, R_d, $\Gamma*$, K_c, and K_o. The proportion of A/C_i sets for which estimation failed altogether, and the range of the estimates and their standard error, are presented in Tables 2 and 3. Hougaard's measure of skewness was computed for all estimated parameters in all cases. The criteria for failure of the procedure were failure to converge, and missing statistics for one or more parameter. In those cases, an estimate may have been produced, but the procedure could not compute a standard error and test of significance. Failure of the procedure was almost always traceable to multicollinearity, as a result of excessive correlation among parameters. Estimates, standard errors, and the proportion of A/C_i sets for which skewness was > 0.15 are for successful sets only.

We first used 231 A/C_i sets, collected in the course of three experiments conducted between 1994 and 2003 at the US Department of Agriculture (USDA)/Agricultural Research Service (ARS) Plant Science Research Unit in Raleigh, North Carolina, on soybean 'Essex' (*Glycine max* 'Essex') (Reid *et al.*, 1998; Booker *et al.*, 2004) and snap bean (*Phaesolus vulgaris*) (S331 line) (Flowers *et al.*, in press) grown in various

Table 2 Effect of varying number of estimated parameters on the quality of the estimates, for 231, 8–12 observation A/C_i sets (see Table 1 for definitions of abbreviations)

Parameters estimated	% failed estimation	% skewness > 0.15, one or more parameters		$V_{c\,max}$	J	R_d	$\Gamma*$	K_c	K_o
$V_{c\,max}$, J, R_d	0	0	Min	42	53	–1			
			Max	239	372	14			
			Min SE	3	2	0			
			Max SE	63	44	9			
			% skewness > 0.15	0	0	0			
$V_{c\,max}$, J, $\Gamma*$	16	60	Min	62	87		3		
			Max	256	358		68		
			Min SE	2	1		1		
			Max SE	35	26		31		
			% skewness > 0.15	22	0		49		
$V_{c\,max}$, J, K_c	30	100	Min	35	97			104	
			Max	1016	336			3538	
			Min SE	2	1			7	
			Max SE	1587	14			5764	
			% skewness > 0.15	100	0			100	
$V_{c\,max}$, J, K_o	51	100	Min	59	122				1
			Max	10591	339				107146
			Min SE	5	1				26
			Max SE	268672	13				25916666
			% skewness > 0.15	100	0				100
$V_{c\,max}$, J, R_d, $\Gamma*$	40	100	Min	47	77	–3	0		
			Max	283	357	18	112		
			Min SE	2	1	0	1		
			Max SE	64	64	20	183		
			% skewness > 0.15	10	11	48	98		
$V_{c\,max}$, J, R_d, K_c	28	100	Min	39	112	0		55	
			Max	6214	373	19		63971	
			Min SE	1	3	1		10	
			Max SE	402718	68	15		4166953	
			% skewness > 0.15	96	86	94		100	
$V_{c\,max}$, J, R_d, K_o	67	100	Min	54	112	0			7
			Max	1909	373	13			20708
			Min SE	13	6	1			48
			Max SE	16867	68	15			3714101
			% skewness > 0.15	100	91	99			100

'% failed estimation' is the proportion of A/C_i sets for which estimation failed to converge, or for which some estimates or statistics could not be computed. Estimation did not necessarily fail for the same sets under all combinations of parameters. '% skewness > 0.15, one or more parameters' is the proportion of sets for which the absolute value of Hougaard's measure of skewness exceeded 0.15. 'Min' and 'Max' are the minimum and maximum estimated values of a parameter, for the sets that did not fail. 'min SE'and 'max SE' are the smallest and largest standard errors for the estimated value of the parameters.

New Phytologist

Table 3 Effect of varying number of estimated parameters on the quality of the estimates, for 32, 22-to-39-observation A/C_i sets (see Table 1 for definitions of abbreviations)

Parameters estimated	% failed estimation	% skewness > 0.15, one or more parameters		$V_{c\,max}$	J	R_d	Γ_*	K_c	K_o
$V_{c\,max}$, J, R_d	22	0	Min	62	136	0			
			Max	198	357	11			
			Min SE	5	7	1			
			Max SE	26	30	4			
			% skewness > 0.15	0	0	0			
$V_{c\,max}$, J, Γ_*	25	72	Min	64	135		5		
			Max	199	338		49		
			Min SE	4	7		5		
			Max SE	16	21		29		
			% skewness > 0.15	8	0		65		
$V_{c\,max}$, J, K_c	22	100	Min	62	164			157	
			Max	1306	326			4293	
			Min SE	9	4			38	
			Max SE	5478	20			18535	
			% skewness > 0.15	100	0			100	
$V_{c\,max}$, J, K_o	28	100	Min	62	164				12
			Max	1305	326				1652
			Min SE	11	4				44
			Max SE	5477	14				7797
			% skewness > 0.15	100	0				100
$V_{c\,max}$, J, R_d, Γ_*	63	100	Min	135	217	–4	6		
			Max	191	366	10	93		
			Min SE	9	8	3	10		
			Max SE	20	34	10	47		
			% skewness > 0.15	17	0	58	100		
$V_{c\,max}$, J, R_d, K_c	28	100	Min	97	175	2		150	
			Max	1283	372	14		5431	
			Min SE	10	11	2		68	
			Max SE	23232	49	10		102204	
			% skewness > 0.15	100	100	100		100	
$V_{c\,max}$, J, R_d, K_o	38	100	Min	97	175	2			9
			Max	1284	372	12			10371
			Min SE	10	11	2			77
			Max SE	23242	49	10			349731
			% skewness > 0.15	100	100	100			100

'% failed estimation' is the proportion of A/C_i sets for which estimation failed to converge, or for which some estimates or statistics could not be computed. Estimation did not necessarily fail for the same sets under all combinations of parameters. '% skewness > 0.15, one or more parameters' is the proportion of sets for which the absolute value of Hougaard's measure of skewness exceeded 0.15. 'Min' and 'Max' are the minimum and maximum estimated values of a parameter, for the sets that did not fail. 'Min SE' and 'Max SE' are the smallest and largest standard errors for the estimated value of the parameters.

atmospheres and at various temperatures. In all sets, A was recorded at nine to 12 levels of C_i, ranging between 26 and 1350 μmol mol^{-1}. Measurements were collected using three different instruments, and included A/C_i sets recorded from plants under stress, and well into the senescence phase. It is important to note that many of these sets were far from the ideal presented in Fig. 2 and, for example, did not show any evidence of distinct segments under visual examination. The aim of including such sets is to avoid using only data that have been chosen for their conformity to model expectations. However, only sets that supported estimation of at least $V_{c\,max}$, J, and R_d were included. In other words, simultaneous estimation using the SAS program included as Appendix S1 succeeded in identifying two segments in all 231 sets, even though two segments were not always obvious from examining the data. Sets that did not fail, but for which one of the parameters did not pass the significance test (t-test with $\alpha = 0.05$), are included also. Even though, strictly speaking, nonsignificant parameters should be removed from the model, many plant physiologists consider the function of the FvCB model as a source of values for photosynthesis parameters paramount, and choose to accept them.

Thirty-two A/C_i sets, each with between 22 and 39 levels of C_i, were then created by aggregating 93 sets corresponding to

multiple experimental subsamples (multiple leaves on the same plant, recorded at the same time). Besides increased dispersion of the data with respect to A, some of the 32 sets were also far from the ideal presented in Fig. 2. Seven of them failed when only $V_{c\,max}$, J, and R_d were included.

Tables 2 and 3 include results from estimation of seven combinations of either three or four parameters. It is immediately apparent that, although the FvCB model is nonlinear by virtue of its segmentation, $V_{c\,max}$, J, and R_d have the same excellent statistical properties as parameters of linear models. This is only true, however, when they are the only parameters to be estimated. As shown by the proportion of A/C_i sets for which skewness exceeded 0.15 for one or more parameters, and by consideration of the magnitude of standard errors relative to parameters, the introduction of any other parameter creates serious difficulties. From combinations of three parameters, it appears that reliable estimation of K_c and K_o from A/C_i data is always problematic, as is, to a somewhat lesser extent, that of $\Gamma*$. Four parameters cannot be estimated reliably, and while estimation succeeded with up to five parameters in some of the 32 sets with increased levels of C_i, adding further parameters can only worsen these problems. The procedure failed with all A/C_i sets when estimation of six parameters was attempted. There was no clear improvement when data included more levels of C_i than customary. This is important, because it strongly suggests that the difficulty in estimating $\Gamma*$, K_c, and K_o from A/C_i data is an intractable limitation of the model.

As pointed out by Long & Bernacchi (2003), once all parameters except $V_{c\,max}$, J, and R_d have been fixed, both A_c and A_j are in fact linear functions. Combining them into a segmented model produces a nonlinear model, but it is possible that having the parameter R_d included in both segments improves the overall estimation properties of the joint model.

These results suggest that, given reliable data, and provided that reliable values are available independently for K_c, K_o, and $\Gamma*$, very good estimates of $V_{c\,max}$ and J can be obtained from gas exchange data through simultaneous estimation. The estimate of R_d produced by simultaneous estimation is expected to be more reliable than that obtained through disjunct estimation, because it is not based only on A_c, and on part of the data. However, its standard error tends to be greater, relative to its estimate, than that of $V_{c\,max}$ or J, and its worth remains diminished by the necessity to fix the value of $\Gamma*$. (Because $C_i = \Gamma*$ at $A = -R_d$, setting $\Gamma*$ imposes a constraint on R_d.)

Internal conductance g_i

Although assessing the reliability of various methods of measuring g_i has proved difficult (Warren, 2006), it is nonetheless very likely that in many cases, possibly most, the assumption that $C_i = C_c$ is untenable. Variation in g_i has recently been shown to correspond to species, leaf age, and various environmental factors (Niinemets *et al.*, 2005, 2006; Warren & Dreyer, 2006). The effect of a nonnegligible resistance on estimates of the parameters of the FvCB model can easily be quantified: C_c values can first be calculated from C_i, using values of g_i that are either hypothetical or measured independently. Parameter estimates can then be obtained and compared using either C_c or C_i as the independent variable. Ethier & Livingston (2004) have done this on a small scale, but they unfortunately used disjunct estimation, with a single, universal cut-off value for subsetting their data.

They also proposed that g_i can be estimated (regressionally) from conventional A/C_i data. Beyond some algebraic operations, the method they describe requires arbitrary subsetting, the exclusion of some data altogether, and the estimation of four parameters, in addition to g_i. Whether or not ordinary A/C_i data are sufficient to support estimation of g_i can be tested by adding Eqn 6 to the FvCB model in order to define C_c from C_i, and fitting the joint model with C_c as the independent variable, instead of C_i. The model thus includes g_i as a parameter to be estimated, in addition to $V_{c\,max}$, J, and R_d. This produces reasonable-seeming estimates of g_i from many A/C_i sets. Estimation may fail when g_i makes no contribution to the model, suggesting that, in those data sets, the assumption that C_i is equal to C_c would be inconsequential. The main difficulty with estimating g_i from a single A/C_i set is similar to that affecting estimation of $\Gamma*$: both parameters relate to variation in the independent variable. Not surprisingly, introducing g_i in the estimation results in values for Hougaard's coefficient of skewness > 0.15 for at least one parameter in all cases. Therefore, while it is quite feasible to estimate g_i from a single A/C_i set, careful validation would be required in the context of prediction, irrespective of the regressional method used. Issues regarding the estimation of g_i require more extensive treatment, and, as indicated in 'Extensions of simultaneous estimation', simultaneous estimation may be a preliminary step in achieving better estimation of g_i.

Local minima and flat regions in the objective function

The objective function of nonlinear models is sometimes marked by the presence of multiple minima, only one of which is the global one, and corresponds to the authentic best estimate of the parameters. Minimization algorithms require that starting values be supplied, from which they will efficiently converge to a minimum. Whenever a local minimum exists in the objective function, and a single combination of starting parameter values is given to the minimization algorithm such that the initial evaluation is closer to the local minimum than to the global one, the likelihood of convergence to the local one is very high (Seber & Wild, 1989; SAS Institute Inc, 2005). Examination of SSE for matrices of fixed values of subsets of the parameters showed the frequent presence of multiple local minima for the FvCB model. Some choices of starting values are therefore very likely to lead to incorrect estimates.

Another difficulty arises when the objective function presents unbounded flat areas, where changing the value of a parameter has no effect, or very little, on the objective function. In that case, the parameter should be eliminated. In the two-segment FvCB model, this does not occur when $V_{c\,max}$, J, and R_d are the only parameters estimated, but g_i is affected by this problem in some data sets.

Estimation problems and solutions

Local minima

The difficulty posed by local minima is readily overcome by choosing starting values for the parameters through a preliminary high-density grid search, and it is imperative that one be used. In a grid search, several starting values are supplied for one or all parameters to be estimated, and an SSE is computed using every combination of those starting values, but without minimization. The combination of starting values that yields the smallest SSE is then used to start the minimization and estimate the model. Use of high grid density, and wide ranges of values in constructing the grid, ensures that the starting combination will be in sufficiently close proximity to the global minimum to avoid convergence to a local one. The programs included as Appendix S1 incorporate a preliminary grid search.

Biologically implausible estimates

Under the assumption that the FvCB model constitutes a good representation of the response of A to varying CO_2, grossly implausible estimates should prompt a close review of the particular data, and of the suitability of the model. If problematic parameter estimates can be held as accidents of normal data variability, as with occasional small negative values of R_d, and do not affect many replicate A/C_i sets in any systematic manner, they should be allowed to stand, lest some bias be introduced. It is important to note that this only applies to contexts in which a large number of A/C_i sets are being used, with the objective of analyzing the effect of some factor of interest on photosynthesis. It does not apply when there are few sets, and the objective is to parametrize a predictive model.

It is possible to prohibit the estimation procedure from reaching impossible values by constraining the range of a parameter (without necessarily setting it to a unique value), but great caution must be exercised in adding *ad hoc* bounds to estimation procedures. Often, the procedure simply converges to the bound value, doing little more than setting the parameter at that value, while obscuring the fact that the parameter is indeed being set. Wide bounds may have some utility in keeping the minimization from drifting into nonsensical ranges, but, for A/C_i sets that do result in a bound being reached, the suitability of the segmented model must be reconsidered, and the quality of the data closely reviewed.

Single-phase data

Regardless of what segments were sampled when measuring A, the procedure will attempt to estimate as many segments as have been requested. If the data being fitted comprise samples of only one segment, the procedure will use one of the component functions to model A over the entire recorded range of C_i, while extending the other segment solely to the extrapolated range, resulting in meaningless estimates for that function or functions. In the process, $C_{i\,tr}$ is placed at one extreme of the range of C_i values actually recorded, or even well outside of it. Depending on the software used, confidence intervals may also be missing for the meaningless parameter(s). The procedure may also fail to converge. This problem can be readily identified by examining $C_{i\,tr}$, and confidence intervals for the parameters. A plot of the data and the predicted curve will also show that the transition point is at the edge, or even outside of the recorded data, and that at least one entire segment does not overlap with any observation. In these cases, a model with fewer segments is obviously more appropriate, and should be used. Data that support estimation of the A_p segment are rare, and adding it to the model will usually lead to this problem, but fitting two segments to data that only support one should also be avoided.

In very rare circumstances, data may include samples of both segments, but random errors result in an A/C_i set for which one of the two functions alone happens to provide a better fit than the two functions combined, despite the underlying presence of two phases. In those rare A/C_i sets, the same symptoms may appear, including extreme values of $C_{i\,tr}$. The presence of this condition can be judged by consideration of A/C_i sets corresponding to closely related experimental units, such as replicates, recorded closely in time. If the values of A in those related sets are similar, and if fitting both segments to them proceeds normally, then it is likely that the isolated problem set does comprise two phases. In these circumstances, constraining $C_{i\,tr}$, for that single set only, becomes justifiable. Values for the constraint are best obtained from the unconstrained fitting of the model to the related sets. Note that, when $C_{i\,tr}$ is constrained, parameters are still estimated from the entirety of the data.

Multicollinearity and excessive correlation between parameters

For some data, any number of different combinations of parameter values all result in optimal fits, because of multicollinearity. Some data may also involve correlation among parameters that exceeds the tolerance of the procedure. Either type of data may result in failure to converge, grossly outsized standard errors, singularity in the Hessian, or estimates identified as biased, all indicative of failure of the procedure. Various statistical software may report different errors for the same underlying problem, which can be viewed

as a special case of overparametrization. Correlation between parameters is largely inherent to the model, as reviewed in 'Disjunct segments estimation methods,' and can be examined by outputting the correlation matrix, or collinearity diagnostics as available. The Levenberg–Marquardt minimization algorithm is reported to have greater tolerance of excessive correlation than other algorithms (Schabenberger *et al.*, 1999; SAS Institute Inc., 2005; D. Dickey, pers. comm.), but reparametrization is generally recommended. In view of the value of the FvCB model as a standard, this option may not be desirable. Assuming that suitable starting values for the estimation have been supplied, the addition of a single mild constraint, such as a constraint on $C_{i\,tr}$, is sufficient to resolve it, and should be preferred.

Extensions of simultaneous estimation

The method outlined herein can be extended to estimating models that require more information than a single A/C_i set. The objectives of including additional information include: modeling of environmental effects on photosynthesis, improving estimates of $V_{c\,max}$, J, and R_d, and estimating more than three parameters.

For example, knowing that theory suggests that values of $\Gamma*$ and R_d at the intersection point ($\Gamma*$, $-R_d$) remain invariant with variable irradiance, we can speculate that separate A/C_i sets obtained at different irradiances might support simultaneous estimation of the various corresponding $V_{c\,max}$ and J, and common $\Gamma*$ and R_d. Preliminary results indicate that estimation of $\Gamma*$ and g_i may be strengthened too. In a further extension, estimation of the FvCB model and of an accepted empirical model of the dependence of A on irradiance may be combined into estimation of a single, higher dimensional model. Finally, simultaneous estimation may be integrated into systems of regression equations, in which estimation is conducted from separate data that do not share variables, but whose models share at least one parameter. A prominent example in photosynthesis research is provided by gas exchange and chlorophyll fluorescence data, both of which yield estimates of J. Systems of regression equations may yield improved estimates of parameters of both models based on gas exchange and models based on chlorophyll fluorescence data.

Conclusion

Segmented regression, which is easily implemented using statistical software that processes conditional syntax, obviates the need for arbitrary determination of transition points and subsetting of the data before analysis, and thus removes bias while improving accuracy and precision. The method lends itself to rapid fitting of the FvCB model to large numbers of A/C_i sets, and dramatically improves the efficiency of processing gas exchange data. Estimates of $V_{c\,max}$, J, and R_d have excellent statistical properties when they are the only parameters to be estimated, but estimation of more than these three parameters from a single, 8–12 observation A/C_i set may be difficult. Because of the prevalence of local minima, implementations that identify starting values for minimization in proximity to the global minimum are indispensable. Some specific difficulties in estimation can be resolved by placing bounds on parameters, or by the carefully considered application of other mild constraints. The methods described may be extended to fitting higher dimensional models, and to systems of regression equations.

Acknowledgements

We thank David Dickey (Department of Statistics, NCSU), Pierre Montpied (Unité Mixte de Recherche INRA-UHP 1137), Kevin Tu (Department of Integrative Biology, UCB), and Patrick B. Morgan (Li-Cor Biosciences, Inc.), for repeated discussions and comments that contributed to this paper in its present form.

References

Booker FL, Fiscus EL, Miller JE. 2004. Combined effects of elevated atmospheric carbon dioxide and ozone on soybean whole-plant water use. *Environmental Management* **33**: S355–S362.

von Caemmerer S. 2000. *Biochemical models of leaf photosynthesis.* Collingwood, Australia: CSIRO Publications.

von Caemmerer S, Farquhar GD. 1981. Some relationships between the biochemistry of photosynthesis and the gas exchange of leaves. *Planta* **153**: 376–387.

Cai T, Dang Q. 2002. Effects of soil temperature on parameters of a coupled photosynthesis–stomatal conductance model. *Tree Physiology* **22**: 819–827.

Dreyer E, Le Roux X, Montpied P, Daudet FA, Masson FA. 2001. Temperature response of leaf photosynthetic capacity in seedlings from seven temperate tree species. *Tree Physiology* **21**: 223–232.

Ethier GJ, Livingston NJ. 2004. On the need to incorporate sensitivity to CO_2 transfer conductance into the Farquhar–von Caemmerer–Berry leaf photosynthesis model. *Plant, Cell & Environment* **27**: 137–153.

Farquhar GD, von Caemmerer S, Berry JA. 1980. A biochemical model of photosynthetic CO_2 carbon dioxide assimilation in leaves of C_3 carbon pathway species. *Planta* **149**: 78–90.

Flexas J, Díaz-Espejo A, Berry JA, Cifre J, Galmés J, Kaldenhoff R, Medrano H, Ribas-Carbó M. 2007. Analysis of leakage in IRGA's leaf chambers of open gas exchange systems: quantification and its effects in photosynthesis parameterization. *Journal of Experimental Botany* **58**: 1533–1543.

Flowers MD, Fiscus EL, Burkey KO, Booker FL, Dubois J-JB. (in press). The effects of ozone on the photosynthesis and yield of snap bean (*Phaesolus vulgaris* L.). *Environmental and Experimental Botany.*

Haines LM, O'Brien TE, Clarke GPY. 2004. Kurtosis and curvature measures for nonlinear regression models. *Statistica Sinica* **14**: 547–570.

Harley PC, Tenhunen JD. 1991. Modeling the photosynthetic response of C_3 leaves to environmental factors. In: Boote KJ, Loomis RS, eds. *Modeling crop photosynthesis – from biochemistry to canopy: Proceedings of a Symposium Sponsored by Division C-2 of the Crop Science Society of America and Division a-3 of the American Society of Agronomy in Anaheim, California, 29 November 1988.* Madison, WI, USA: Crop Science Society of America: American Society of Agronomy, 17–39.

Leuning R. 1997. Scaling to a common temperature improves the correlation between the photosynthesis parameters J_{max} and $V_{c\,max}$. *Journal of Experimental Botany* **48**: 345–347.

Long SP, Bernacchi CJ. 2003. Gas exchange measurements, what can they tell us about the underlying limitations to photosynthesis? Procedures and sources of error. *Journal of Experimental Botany* **54**: 2393–2401.

Manter DK, Kerrigan J. 2004. *A*/C_i curve analysis across a range of woody plant species: influence of regression analysis parameters and mesophyll conductance. *Journal of Experimental Botany* **55**: 2581–2588.

Medlyn BE, Dreyer E, Ellsworth D, Forstreuter M, Harley PC, Kirschbaum MUF, Le Roux X, Montpied P, Strassemeyer J, Walcroft A, Wang K, Loustau D. 2002. Temperature response of parameters of a biochemically based model of photosynthesis. II. A review of experimental data. *Plant, Cell & Environment* **25**: 1167–1179.

Niinemets U, Cescatti A, Rodeghiero M, Tosens T. 2005. Leaf internal diffusion conductance limits photosynthesis more strongly in older leaves of Mediterranean evergreen broad-leaved species. *Plant, Cell & Environment* **28**: 1552–1566.

Niinemets U, Cescatti A, Rodeghiero M, Tosens T. 2006. Complex adjustments of photosynthetic potentials and internal diffusion conductance to current and previous light availabilities and leaf age in Mediterranean evergreen species *Quercus ilex*. *Plant, Cell & Environment* **29**: 1159–1178.

Nippert JB, Fay PA, Knapp AK. 2007. Photosynthetic traits in C_3 and C_4 grassland species in mesocosm and field environments. *Environmental and Experimental Botany* **60**: 412–420.

Onoda Y, Hikosaka K, Hirose T. 2005. The balance between RuBP carboxylation and RuBP regeneration: a mechanism underlying the interspecific variation in acclimation of photosynthesis to seasonal change in temperature. *Functional Plant Biology* **32**: 903–910.

Ratkowsky DA. 1990. *Handbook of nonlinear regression models*. New York, NY, USA: M. Dekker.

Rawlings JO, Pantula SG, Dickey DA. 1998. *Applied regression analysis. A research tool*. New York, NY, USA: Springer.

Reid CD, Fiscus EL, Burkey KO. 1998. Combined effects of chronic ozone and elevated CO_2 on Rubisco activity and leaf components in soybean (*Glycine max*). *Journal of Experimental Botany* **49**: 199–2011.

SAS Institute Inc. 2005. *SAS 9 1 3. Help and documentation*. Cary, NC, USA: SAS Institute Inc.

Schabenberger O, Tharp BE, Kells JJ, Penner D. 1999. Statistical tests for hormesis and effective dosages in herbicide dose–response. *Agronomy Journal* **91**: 713–721.

Schultz HR. 2003. Extension of a Farquhar model for limitations of leaf photosynthesis induced by light environment, phenology and leaf age in grapevines (*Vitis vinifera* L. cvv. White Riesling and Zinfandel). *Functional Plant Biology* **30**: 673–687.

Seber GAF, Wild CJ. 1989. *Nonlinear regression*. New York, NY, USA: Wiley.

Warren CR. 2006. Estimating the internal conductance to CO_2 movement. *Functional Plant Biology* **33**: 431–442.

Warren CR, Dreyer E. 2006. Temperature response of photosynthesis and internal conductance to CO_2: results from two independent approaches. *Journal of Experimental Botany* **57**: 3057–3067.

Wohlfahrt G, Bahn M, Cernusca A. 1999. The use of the ratio between the photosynthesis parameters P_{ml} and $V_{c\,max}$ for scaling up photosynthesis of C3 plants from leaves to canopies: a critical examination of different modelling approaches. *Journal of Theoretical Biology* **200**: 163–181.

Wullschleger SD. 1993. Biochemical limitations to carbon assimilation in C_3 plants – A retrospective analysis of the A/C_i curves from 109 species. *Journal of Experimental Botany* **44**: 907–920.

Supplementary Material

The following supplementary material is available for this article online.

Appendix S1 SAS programs for estimation of the FvCB model, with instructions.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1469-8137.2007.02182.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the journal at *New Phytologist* Central Office.

